JOURNAL OF INFORMATION PROCESSING SYSTEMS JIPS

The Improved Joint Bayesian Method for Person Re-identification Across Different Camera

Ligang Hou*, Yingqiang Guo*, and Jiangtao Cao*

Abstract

Due to the view point, illumination, personal gait and other background situation, person re-identification across cameras has been a challenging task in video surveillance area. In order to address the problem, a novel method called Joint Bayesian across different cameras for person re-identification (JBR) is proposed. Motivated by the superior measurement ability of Joint Bayesian, a set of Joint Bayesian matrices is obtained by learning with different camera pairs. With the global Joint Bayesian matrix, the proposed method combines the characteristics of multi-camera shooting and person re-identification. Then this method can improve the calculation precision of the similarity between two individuals by learning the transition between two cameras. For investigating the proposed method, it is implemented on two compare large-scale re-ID datasets, the Market-1501 and DukeMTMC-reID. The RANK-1 accuracy significantly increases about 3% and 4%, and the maximum a posterior (MAP) improves about 1% and 4%, respectively.

Keywords

Joint Bayesian, Multi-Camera Shooting, Person Re-identification, Superior Measurement

1. Introduction

The re-identification of individuals across different cameras has fascinated many researchers in the area of computer vision. This technique is mainly used in criminal investigation and image retrieval, however, certain problems still remain to be addressed to effectively re-identify an individual using this approach. These problems include differences in camera position and performance, varying illumination conditions, variability of the individuals themselves, and varying appearances of different backgrounds.

Usually, researchers use standard distances, such as the Mahalanobis distance [1], the Euclidean distance [2], and the Bhattacharyya distance [3], to measure the similarity between two identities. However, when the person with the same ID crosses multiple non-overlapping cameras, different appearance features are affected by different factors such as the angle of view and the illumination.

Four images of the Market-1501 dataset [4] are shown in Fig. 1. The images have the same label and are shoot by disjoint cameras. From the chart, the differences which include the difference of color, contour, texture difference and background are shown. These different points bring great complexity of the person re-identification task.

Fig. 1. Pictures of individual cross disparate camera with the same label on Market-1501 dataset [4].

There are two ways to address these problems. First, features are designed and selected according to the appearance individual of illumination changes, and the human body deformation robust. Kviatkovsky et al. [5] presented a colour structure of an internal distribution that is invariant to illumination and a covariance descriptor that is used for pedestrian re-identification. This unsupervised learning method is used to extract significant features of human images, which, in turn, can be used for pedestrian re-identification [6]. Second, in this approach, distance metric learning is used to learn a feature transformation or distance measure by which the subject is at the minimum distance from different cameras, while the other people are at the maximum distance from the same camera [7]. The distance learning metric mechanism is optimised so that the distance between the correct matches is less than the error matching and the probability of the distance between the two matches is maximised. A locally adaptive decision function consists of a joint-distance measure model and the locally adaptive threshold rule is proposed to pedestrian re-identification and to ensure that a good recognition rate is achieved [8]. Liao et al. [9] proposed the LOMO feature, which analyses the horizontal occurrence of local features and maximises the occurrence to make a stable representation to viewpoint changes. Similarly, Zheng et al. [10] used metric learning as well, but formulated it in a probabilistic manner. These researchers sought a distance that maximises the probability of the distance between a matching pair being smaller than that between a non-matching pair. Some new researches used the combined joint Bayesian methods to improve the recognition accuracy, but without considering the changes between different cameras [11-13].

Most of the above-mentioned methods are manual methods and do not use depth neural networks to extract features; therefore, these methods have some shortcomings. To overcome these disadvantages, a more efficient approach that can exploit the natural constraints given by the person re-identification task is proposed and constitutes the contribution to this research. First, the features are extracted from a network with potential camera distributions. Second, the Joint Bayesian formulation is redefined that can better serve person re-identification across different camera distributions. Then, the improved Joint Bayesian method is built by combining the characteristics of multi-camera shooting and person re-identification with the Joint-Bayesian method. The proposed method is better adapted to the scene re-ID by learning the same pedestrian lighting and background changes under different cameras.

This paper is organised as follows. In Section 2, the Joint-Bayesian method, which has been implemented for Bayesian face recognition, is introduced. In Section 3, the improved JBR method of related re-identification process analysis is proposed. In Section 4, the implementation and comparison of the existing and proposed methods are presented on two large-scale person re-ID datasets, that is, Market-1501 [4] and DukeMTMC-reID [14]. In Section 5, the concluding remarks are given.

2. Represented Joint Bayesian Method

A brief description of the research results of the former researchers is represented in this section. Based on the previous Bayesian face recognition method [15], the verification task can be formulated as a Bayesian decision problem. The appearance of a face is mainly influenced by two factors: intra-personal variation and extra-personal variation. Here two pictures, which can be faces or other object figures, are represented as x_1 and x_2. Δ represents the difference between the two pictures. Let H_I and H_E represent the intra-personal hypothesis and the extra-personal hypothesis. Based on the maximum a posterior (MAP) rule, the decision on the verification task is made by testing a log likelihood ratio $r(\Delta)$.

$$r(\Delta) = \log \frac{P(\Delta|H_I)}{P(\Delta|H_E)} \tag{1}$$

This ratio r can be used to measure the similarity between two pictures. Chen et al. [15] proposed a joint formulation to directly model the joint distribution of $\{x_1, x_2\}$. A verified object can be represented by the sum of two independent Gaussian variables as follows:

$$x = \mu + \xi \tag{2}$$

where x is the observed object with the mean of all objects subtracted, μ represents its identity, and ξ is the object variation (e.g., lightings, poses) within the same identity. μ and ξ follow two Gaussian distributions, $N(0, S_\mu)$ and $N(0, S_\xi)$. Where S_μ and S_ξ are two unknown covariance matrices.

Based on the above prior knowledge, we can obtain a Gauss joint distribution with a mean value of 0. Both μ and ξ are independent, and then the covariance of two face features is obtained as follows:

$$cov(x_1, x_2) = cov(\mu_1, \mu_2) + cov(\xi_1, \xi_1) \tag{3}$$

For the same person, μ_1 and μ_2 are the same, ξ_1 and ξ_2 are independent. Thus, the covariance matrix of the $P(x_1, x_2 \mid H_I)$ distribution can be calculated as follows:

$$\Sigma_I = \begin{bmatrix} S_\mu + S_\xi & S_\mu \\ S_\mu & S_\mu + S_\xi \end{bmatrix} \tag{4}$$

Based on the premise that people are different from each other, both μ and ξ are independent. Thus, the covariance matrix of the $P(x_1, x_2 \mid H_E)$ distribution can be calculated as follows:

$$\Sigma_E = \begin{bmatrix} S_\mu + S_\xi & 0 \\ 0 & S_\mu + S_\xi \end{bmatrix} \tag{5}$$

With the two conditional joint probabilities, the log likelihood ratio $r(x_1, x_2)$ can be obtained in a closed form with simple algebra operations:

$$r(x_1, x_2) = \log \frac{P(x_1, x_2 \mid H_I)}{P(x_1, x_2 \mid H_E)} = x_1^T A x_1 + x_2^T A x_2 - 2 x_1^T G x_2 \tag{6}$$

where

$$A = (S_\mu + S_\xi)^{-1} - (F + G)$$

$$\begin{pmatrix} F + G & G \\ G & F + G \end{pmatrix} = \begin{pmatrix} S_\mu + S_\xi & S_\mu \\ S_\mu & S_\mu + S_\xi \end{pmatrix}^{-1} \tag{7}$$

Both A and G are negative semi-definite matrices. If A = G, the negative log likelihood ratio will degrade to the Mahalanobis distance. The log likelihood ratio metric is invariant to any full rank linear transform from the feature. More details of the above formulations can be found in the study of Chen et al. [15].

3. Improved Joint Bayesian Method

In this section, to clearly demonstrate the proposed method, the structure of improved Joint Bayesian method is shown in Fig. 2. Fig. 2 shows the images (left) of some pedestrian and the features (right, red filled circles) extracted from the convolutional neural network regarding the distribution of the cameras. C_i indicates the label of the camera, where $i=1, \ldots, 6$. Different colours with double-headed arrows represent distance measurements with different Joint Bayesian matrices that are trained by features of each pair of cameras. For example, the green double-headed arrow shown in Fig. 2 indicates that if we calculate the feature of C_6 to the feature of C_3 of the distance, it is necessary to select the Joint-Bayesian matrices that are trained by features from the C_6 and C_3 cameras. The procedure is to extract features from a photo that is distributed across different cameras and then calculate the distance between these features using the Joint-Bayesian matrix.
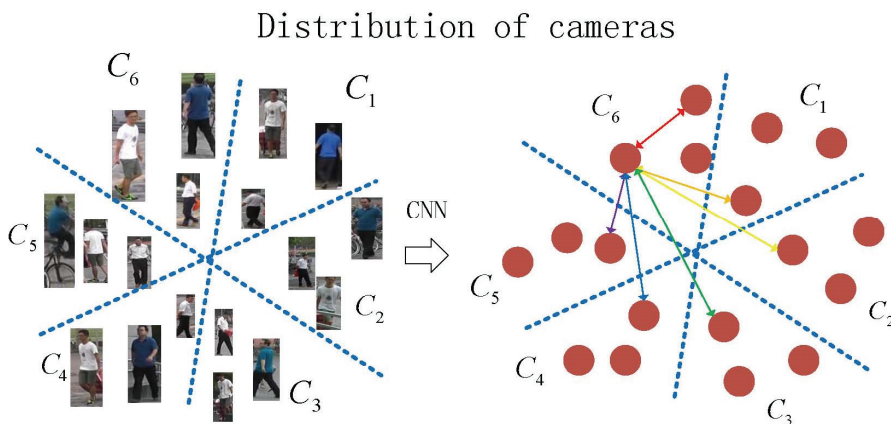

Fig. 2. Images (left) of some pedestrian and features (right, red filled circles) of the cameras' distribution that has been extracted from the convolutional neural network. C_i indicates the label of each camera. Different colours with double-headed arrows represent distance measurements.
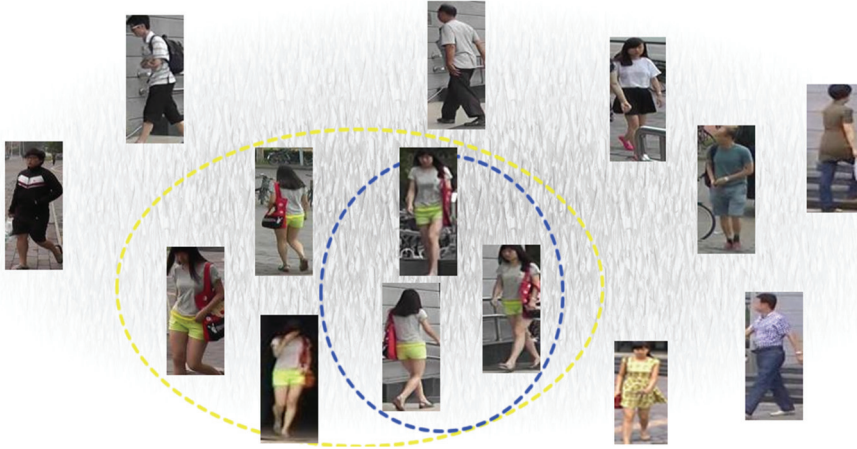
Fig. 3. Distribution of the entire region of the representative person, in which the yellow dotted line represents the distribution of the same person, and the blue dotted line represents the same camera under the same label, modelled by Gaussians.

Similar to face recognition, the picture of a pedestrian consists of two parts: differences between different individuals and differences between images of the same individual under varying conditions. The distribution of the entire region of the representative person, in which the yellow dotted line represents the distribution of the same person and the blue dotted line represents the same camera under the same label, is shown in Fig. 3. In the person re-identification task, the person distribution includes not only the distribution of different people in formulation (6) and the distribution of people under varying conditions, but also the feature distribution of the same person under disjoint cameras. When the person characteristic distance is calculated, in most cases, we need to calculate the similarity of two features under different cameras. Thus, the log likelihood ratio of distance can be redefined by camera pairs as follows:

$$S_{pairs}(\Delta) = \log \frac{P(\Delta|H_{Ipairs})}{P(\Delta|H_{Epairs})} \tag{8}$$

In the above formula, $S_{pairs}(\Delta)$ denotes the similarity between two features that were taken from a pair of cameras, $P(\Delta|H_{Ipairs})$ denotes the likelihood ratio of the distance of the two features that captured the pair of cameras in the same individual condition, and $P(\Delta|H_{Epairs})$ denotes the likelihood ratio under varying individual conditions. To learn the distribution of different cameras, formula (6) is reorganized as the following formulation:

$$r_{pair}(x_1, x_2) = x_1^T A_{ij} x_1 + x_2^T A_{ij} x_2 - 2x_1^T G_{ij} x_2 \tag{9}$$

A_{ij} and G_{ij}, where i and j represent the camera transfer matrix from the i camera to the j camera. $r_{pair}(x_1, x_2)$ denotes the similarity of two individuals from a pair of cameras. However, it is clear that some of the information is lost and is not used to train the matrix. Because some of the pictures that are included in the whole dataset are captured by a pair of cameras, a global distance formulation is rebuilt as below:

$$r(x_1, x_2) = \alpha r_{global}(x_1, x_2) + (1-\alpha) r_{pair}(x_1, x_2)$$
$$= x_1^T A_t x_1 + x_2^T A_t x_2 - 2 x_1^T G_t x_2 \qquad (10)$$
$$where: A_t = \alpha A + (1-\alpha) A_{ij}$$
$$G_t = \alpha G + (1-\alpha) G_{ij}$$

where $r_{global}(x_1, x_2)$ is the same as in formula (6), and denotes the global distance between two pictures. A and G represent the global Joint Bayesian matrix, and α represents the coefficient of the global and the local Joint Bayesian matrix. With the formula (10), we can calculate the similarity of two images that contain global information and local information.

- Training: The training strategy is similar to that in [15]. First of all, the dimension of features of 256 is reduced to simplify the computational complexity. Then, the global Joint Bayesian matrix is trained by identity labels, without adding camera labels, and the local Joint Bayesian matrices A_{ij} and G_{ij} are applied to the camera label for training, until convergence occurs.
- Testing: During the test period, the application of formulation (6) also needs two features: the reduction of the dimension of 256 and the matrix learned in the training phase. The matrix includes the global matrices (A and G) and the local matrices (A_{ij} and G_{ij}) and is decided based on which camera the two images are taken with. To calculate the distance between two features using formulation (6), the matrix and features must be obtained.

4. Evaluating Results

4.1 Datasets and Settings

The proposed method is tested on Market-1501 and DukeMTMC-reID [14], which are two benchmark datasets for person recognition.

4.1.1 Market-1501

This is a person re-identification labelled dataset. This dataset contains a shot from six cameras of 32,668 labelled images of 1,501 identities. It uses the DPM algorithm to capture the pictures. In these images, training data includes 12,936 pictures and 751 identities, and test data include 23,100 pictures and 751 identities. In the test phase, 3,368 pictures with 750 identities are contained in the query, and 19,732 pictures with 751 labels are contained in the gallery.

4.1.2 DukeMTMC-reID

This dataset contains 1,812 identities captured by 8 cameras. In the dataset, 1,404 identities appear in 8 cameras, and the remaining 408 identities are the distracted images. The training and testing sets have 702 identities. In total, there are 2,228 query images, 16,522 training images, and 17,661 gallery images.

4.2 Experiments on Market-1501

4.2.1 Comparison with the state-of-the-art

One experiment is implemented on Market-1501, and the Siamese network on ResNet-50 is used [4].

The results of this experiment are shown in Table 1. When the Euclidean is used to evaluate the distance, the recognition rate for Rank-1 is 82.21%, and MAP is 62.35% on the Siamese network. When the Joint Bayesian is used to measure the similarity of features with the global Joint Bayesian [15], the recognition rate for Rank-1 is 84.80%, and MAP is 75.77%. For Rank-1, the recognition rate increases by 2.5%, and for MAP, the recognition rate increases by 13.5%. Thus, all these factors show improvement. From these results, it is clearly shown that the JBR method not only works well for face recognition but also shows a better recognition rate for pedestrian recognition tasks. Once the re-ranking algorithm is integrated, by combining with the Joint Bayesian method for obtaining the distance between person features, the recognition rates are also improved. For Rank-1, the recognition rate is 85.90%, which is an increase of 1.1%. For MAP, the recognition rate is 77.36%, which is an increase of 1.59%. Based on these test results, the proposed method is evaluated. Let $\alpha = 0$; for Rank-1, the recognition rate is 85.51%; for MAP, the recognition rate is 66.06%. When the re-ranking algorithm is combined, for Rank-1, the recognition rate is 88.09%, which is the best one. Simultaneously, for MAP, the best recognition rate is obtained, that is, 78.24%. From these evaluation results, it can be seen that the proposed method shows the best recognition performance for pedestrian recognition tasks. The recent research by Wang et al. [11] used the combined Joint Bayesian method and the experiment results were shown in Table 1. The accuracy of pedestrian recognition was improved. Clearly, the proposed method obtains better results. The main reason is that the proposed method takes into account the factors from the perspective of distance measurement and the changes between different cameras.

Table 1. Recognition rates (%) on both the Market-1501 and DukeMTMC-reID datasets

Method	Market-1501					DukeMTMC-reID				
	MAP	Rank-1	Rank-5	Rank-10	Rank-20	MAP	Rank-1	Rank-5	Rank-10	Rank-20
S+E	62.35	82.21	92.90	95.64	97.48	54.03	72.17	85.37	89.32	92.15
S+J	75.77	84.83	92.37	94.78	96.80	53.19	73.65	85.73	89.54	92.32
S+J+R	77.36	85.90	93.07	95.22	97.17	70.16	77.51	86.80	89.81	92.82
S+JBR	66.06	85.51	**94.00**	**96.17**	**97.74**	55.44	76.35	**88.20**	**91.25**	**93.36**
S+JBR+R	**78.24**	**88.09**	92.96	94.27	96.14	**70.91**	**80.07**	87.79	90.17	92.68

The best scores are bold.
S=Siamese network, E=Euclidean distance, J=Joint Bayesian, R=re-ranking, JBR=proposed method of Joint-Bayesian across different cameras for person re-identification.

4.2.2 Results between camera pairs

As shown in Fig. 4, the results between all camera pairs are compared between the Siamese network [4] and the proposed method. Although camera_6 is a 720×576 camera and captures distinct background with other HD cameras, the re-identification accuracy result by the proposed method is more competitive. Thus, the proposed method improves the performance of multi-camera pedestrian recognition.

4.2.3 Effect of α on experiments

To clearly test the effect of α in formulation (10), additional experiments with different α values are given on Market-1501. The results are shown in Fig. 5. It can be seen that changing α directly affects the recognition performance of Rank-1 and MAP. It is shown that the smaller α, the greater Rank-1 and MAP; moreover, when $\alpha = 0$, Rank-1 and MAP approach the maximum.
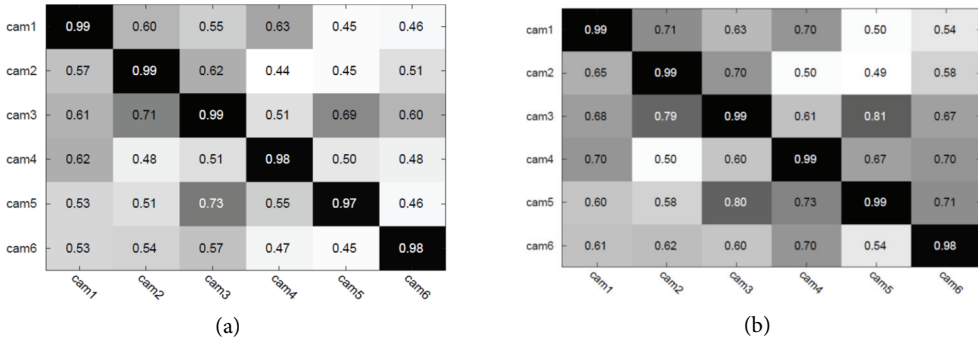
Fig. 4. Confusion matrix between camera pairs for Market-1501: (a) the Siamese network [4] and (b) our proposed method.
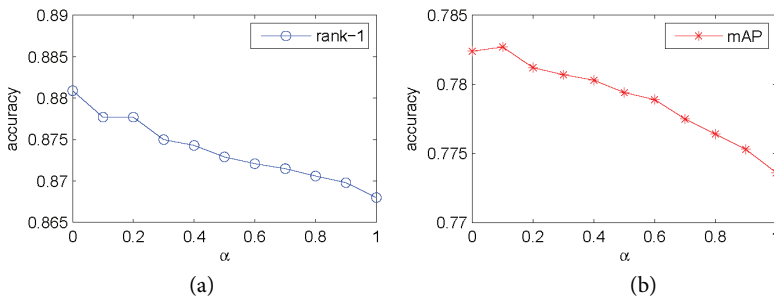

Fig. 5. Recognition results with different α values for Rank-1 (a) and MAP (b).

4.2.4 Ranklist

The performance is compared in a multi-camera scenario under the Joint Bayesian and our algorithm in Fig. 6. The query consists of a pedestrian's picture that needs to be tested. Ranklist is the 8-image set with the highest score in the current query (on the left, the score is the highest and gradually reduces to the right); the pink box represents the wrong result. Below each picture is the pedestrian digital camera number. From Fig. 6, it can be seen that the proposed algorithm has better pedestrian recognition under different cameras and can better deal with this multi-scene situation.
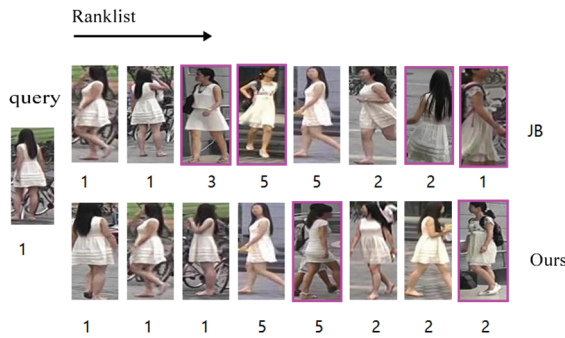

Fig. 6. Comparison of performance in a multi-camera situation under the JB and proposed method. The query consists of a pedestrian picture that needs to be tested. Ranklist is the 8-image set of the pedestrian with the highest score in the current query, the digital camera number is below each picture, and the pink box represents the wrong result.

Finally, in order to demonstrate the performance of the proposed method, the comparison results with other algorithms on Rank-1 and MAP are given in Table 2. The methods named APR and TriNet are selected from [16,17], these are all well-applied methods in the state-of-the-art algorithms and obtain higher recognition rates on Market-1501. Based on these results, it can be seen that the proposed method shows the highest recognition rate on Rank-1.

Table 2. Recognition rates of selected methods on Market-1501

Method	Recognition rate (%)	
	MAP	Rank-1
APR [16]	64.67	84.29
TriNet [17]	69.14	84.92
TriNet+R	81.07	86.67
S+J	75.77	84.83
S+J+R	77.36	85.90
S+JBR	66.06	85.51
S+JBR+R	**78.24**	**88.09**

The best scores are bold.
APR=attribute person recognition, R=re-ranking, S=Siamese network, J=Joint Bayesian, JBR=proposed method of Joint-Bayesian across different cameras for person re-identification.

4.3 Experiments on DukeMTMC-reID

The proposed method is evaluated in DukeMTMC-reID, which has 8 cameras. The test results are shown in Table 3. With the proposed method in Section 3, the Joint Bayesian matrix is trained based on the features from different cameras. The Siamese network is selected as the network architecture that has proven to be the better option to extract the features of individuals [4].

Table 3. Selected best recognition rates on DukeMTMC-reID

Method	Recognition rate (%)	
	MAP	Rank-1
BoW+KISSME [18]	12.17	25.13
LOMO+CQDA [9]	17.04	30.75
Baseline [19]	44.99	65.22
Baseline+LSRO [20]	47.13	67.68
APR [16]	51.88	70.69
SVDNet [21]	56.80	76.70
S+J	53.19	73.65
S+J+R	70.16	77.51
S+JBR	55.44	76.35
S+JBR+R	**70.91**	**80.07**

The best scores are bold.
BoW=bag-of-word, LOMO=local maximal occurrence, CQDA=cross-view quadratic discriminant analysis, LSRO=label smoothing regularization for outliers, APR=attribute person recognition, S=Siamese network, J=Joint Bayesian, R=re-ranking, JBR=proposed method of Joint-Bayesian across different cameras for person re-identification.

To further test the performance of the proposed method, similar to Section 4.2, the other four strategies are built to evaluate the recognition rate in the same dataset. Also, the tests are carried out in MAP and Rank-1, Rank-5, Rank-10, and Rank-20. All the recognition rates are shown in Table 1. Based on the results, the highest recognition rate is obtained with the proposed method, either for MAP or for Rank-1. For Rank-5, Rank-10, and Rank-20, the method without re-ranking obtains better recognition performance.

Also, more comparison results are obtained by using other state-of-the-art algorithms on Rank-1 and MAP, as shown in Table 3. For the other six methods, the results show that SVDNet obtains the best recognition results. That is, the recognition rate is 56.8% for MAP and 76.7% for Rank-1. However, these recognition rates are lower than those of the proposed method, which is 70.91% for MAP and 80.07% for Rank-1. The results clearly show that the proposed method significantly improves the recognition rates and obtains the best performance for pedestrian recognition tasks.

5. Conclusion

A novel method based on the Joint Bayesian method is proposed and implemented to solve the person re-identification under the situation of the same person viewed through multiple cameras. The evaluation results showed that the proposed method achieved better results. Furthermore, the proposed method can be extended to the multi-scene image retrieval field. In this regard, the relevant experiments will be conducted as part of future studies. The limitation of this study is that the computing cost is higher, especially when retrieving a large number of pedestrian features. In the future, it will be continue to study the existence of end-to-end with the Bayesian method, which can be used to identify different persons in the camera scene and to solve the problem of training the Joint Bayesian matrix separately. Meanwhile, it aims to reduce the complexity of the experiment and to enhance the relevance between global and local issue.

Acknowledgement

References

[1] R. De Maesschalck, D. Jouan-Rimbaud, and D. L. Massart, "The mahalanobis distance," *Chemometrics and Intelligent Laboratory Systems*, vol. 50, no. 1, pp. 1-18, 2000.

[2] C. Gosling, "Encyclopedia of distances," *Reference Reviews*, vol. 24, no. 6, pp. 34-34, 2010.

[3] A. Bhattacharyya, "On a measure of divergence between two multinomial populations," *Sankhyā: The Indian Journal of Statistics*, vol. 7, no. 4, pp. 401-406, 1946.

[4] Z. Zheng, L. Zheng, and Y. Yang, "A discriminatively learned CNN embedding for person re-identification," 2017 [Online]. Available: https://arxiv.org/pdf/1611.05666.pdf.

[5]   I. Kviatkovsky, A. Adam, and E. Rivlin, "Color invariants for person reidentification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 7, pp. 1622-1634, 2012.

[6]   R. Zhao, W. Ouyang, and X. Wang, "Unsupervised salience learning for person re-identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Portland, OR, 2013, pp. 3586-3593.

[7]   Z. Li, S. Chang, F. Liang, T. S. Huang, L. Cao, and J. R. Smith, "Learning locally-adaptive decision functions for person verification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Portland, OR, 2013, pp. 3610-3617.

[8]   H. Liu, X. Lv, T. Zhu, and X. Li, "An adaptive feature-fusion method for object matching over non-overlapped scenes," *Journal of Signal Processing Systems*, vol. 76, no. 1, pp. 77-89, 2014.

[9]   S. Liao, Y. Hu, X. Zhu, and S. Z. Li, "Person re-identification by local maximal occurrence representation and metric learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, 2015, pp. 2197-2206.

[10] W. S. Zheng, S. Gong, and T. Xiang, "Person re-identification by probabilistic relative distance comparison," in in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Colorado Springs, CO, 2011, pp. 649-656.

[11] S. Wang, L. Duan, N. Yang, and J. Dong, "Person re-identification with deep dense feature representation and Joint Bayesian," in *Proceedings of 2017 IEEE International Conference on Image Processing (ICIP)*, Beijing, China, 2017, pp. 3560-3564.

[12] Y. C. Chen, X. Zhu, W. S. Zheng, and J. H. Lai, "Person re-identification by camera correlation aware feature augmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 2, pp. 392-408, 2017.

[13] A. Zheng, X. Zhang, B. Jiang, B. Luo, and C. Li, "A subspace learning approach to multishot person reidentification," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2018; http://doi.org/10.1109/TSMC.2017.2784356.

[14] W. Li, R. Zhao, and X. Wang, "Human reidentification with transferred metric learning," in *Computer Vision-ACCV 2012*. Heidelberg: Springer, 2012, pp. 31-44.

[15] D. Chen, X. Cao, L. Wang, F. Wen, and J. Sun, "Bayesian face revisited: a joint formulation," in *Computer Vision-ECCV 2012*. Heidelberg: Springer, 2012, pp. 566-579.

[16] Y. Lin, L. Zheng, Z. Zheng, Y. Wu, Z. Hu, C. Yan, and Y. Yang, "Improving person re-identification by attribute and identity learning," 2017 [Online]. Available: https://arxiv.org/pdf/1703.07220.pdf.

[17] A. Hermans, L. Beyer, and B. Leibe, "In defense of the triplet loss for person re-identification," 2017 [Online]. Available: https://arxiv.org/pdf/1703.07737.pdf.

[18] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: a benchmark," in *Proceedings of the IEEE International Conference on Computer Vision*, Santiago, Chile, 2015, pp. 1116-1124.

[19] L. Zheng, Y. Yang, and A. G. Hauptmann, "Person re-identification: past, present and future," 2016 [Online]. Available: https://arxiv.org/pdf/1610.02984.pdf.

[20] Z. Zheng, L. Zheng, and Y. Yang, "Unlabeled samples generated by GAN improve the person re-identification baseline in vitro," in *Proceedings of the IEEE International Conference on Computer Vision*, Venice, Italy, 2017, pp. 3754-3762.

[21] Y. Sun, L. Zheng, W. Deng, and S. Wang, "SVDNet for pedestrian retrieval," in *Proceedings of the IEEE International Conference on Computer Vision*, Venice, Italy, 2017, pp. 3800-3808.

Ligang Hou https://orcid.org/0000-0001-8508-1310

He is a Professor in School of Information and Control Engineering, Liaoning Shihua University. He received M.S. degree in School of Computer Science from Dalian University of Technology in 1993. His current research interests include intelligent control system and computer vision.

Yingqiang Guo https://orcid.org/0000-0002-1300-0603
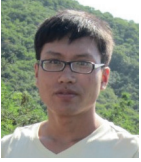
He received B.E. degree in School of Mechanical and Electronic Engineering from Hebei Normal University of Science & Technology in 2015. He is currently a third-year ME student at School of Information and Control Engineering, Liaoning Shihua University. His current research interests include person re-identification and tracking.

Jiangtao Cao https://orcid.org/0000-0002-3830-5753

He is a Professor in School of Information and Control Engineering, Liaoning Shihua University. He received Ph.D. degree in University of Portsmouth in 2009. His current research interests include intelligent system and pattern recognition.